

Cerinte Laborator 2

15 puncte

- Identificarea setului de date (daca e cazul)
- Colectarea setului de date (daca e cazul)
- Intelegerea setului de date (daca e cazul)
- **Analiza exploratorie a datelor folosite ca input pentru proiect**
(mai multe detalii aici: [2023 Analiza Exploratorie a datelor.pdf](#))
 - **Description of data**
 - results from the *describe()* function from python
 - **Handling missing data**
 - drop NULL or missing values
 - fill Missing Values
 - predict Missing values with an ML Algorithm
 - **Handling outliers**
 - BoxPlot
 - Scatterplot
 - Z-score
 - IQR(Inter-Quartile Range)
 - **Understanding relationships and new insights through plots**
 - histogram
 - heatmap
- **Preprocesarea setului de date**
(mai multe detalii aici: [2023 Preprocesarea datelor.pdf](#))
 - **Standardization**
 - StandardScaler
 - MinMaxScaler
 - RobustScaler
 - **Normalization** (L1 or L2)
 - **Discretization**
 - **Encoding categorical features**
 - **Features engineering**
 - Features selection
 - Features extraction

Fiecare echipa va alege *cateva metode* din setul de mai sus pe care le va aplica asupra setului de date corespunzator proiectului echipei. Un fisier care sa contina *raspunsurile* la cerintele Lab 2: *grafice, tabele, etc.* si *explicatii* referitor la rezultatele metodelor aplicate asupra setului propriu de date, dar si un *link inspre codul sursa* care arata aplicarea metodelor alese si rezultatele obtinute (sau o arhiva daca s-a lucrat local) va trebui incarcat in Assignment_L_02.